UCRL-JRNL-147775

Measurements of glottal structure dynamics

John F. Holzrichter*

Lawrence Livermore National Laboratory, Livermore, CA and

Dept. of Applied Science, UC Davis, Davis, CA

Lawrence C. Ng, Gerry J. Burke, Nathan J. Champagne II,

Jeffrey S. Kallman, & Robert M.Sharpe,

Lawrence Livermore National Laboratory

Livermore, CA

James B. Kobler and Robert E. Hillman

Voice and Speech Laboratory, Massachusetts Eye and Ear Infirmary

Boston, MA

John J. Rosowski

Easton Peabody Laboratory, Massachusetts Eye and Ear Infirmary

Boston, MA

This is a Lawrence Livermore National Laboratory, University of California Report
UCRL-JRNL-147775

Abstract:

Low power, radar-like EM wave sensors, operating in a homodyne interferometric mode, are being used to measure tissue motions in the human vocal tract during speech. However, when these and similar sensors are used in front of the laryngeal region during voiced speech, there remains an uncertainty regarding the contributions to the sensor signal from vocal fold movements versus those from pressure induced trachea-wall movements. Several signal-source hypotheses are tested by performing experiments with a subject who had undergone tracheostomy, and who still was able to phonate when her stoma was covered (e.g., with a plastic plate). Laser-doppler motion-measurements of the subject's posterior trachea show small tissue movements, about 15 microns, that do not contribute significantly to signals from presently used EM sensors. However, signals from the anterior wall do contribute. EM sensor and air-pressure measurements, together with 3-D EM wave simulations, show that EM sensors measure movements of the vocal folds very well. The simulations show a surprisingly effective guiding of EM waves across the vocal fold membrane, which, upon glottal opening, are interrupted and reflected. These measurements are important for EM sensor applications to speech signal de-noising, vocoding, speech recognition, and diagnostics.

PACs numbers: 43.70 , 43.70.Aj , 43.70.Jt

I. Introduction:

The use of EM waves for the measurement of human organ positions and motions has been limited because of their coarse transverse resolution which is approximately the wavelength in human tissue, e.g. 2-3 cm for 2Ghz waves. In addition, there is often a longitudinal positional ambiguity regarding the reflection location due to the many-wave long EM wave trains transmitted by interferometric sensors. Impulse EM sensors are being developed to resolve some of these problems, but they are expensive, complicated, and not considered in this paper. However for identified interfaces, requiring non-invasive data, interferometric sensors measure the longitudinal positions to parts in 10^4giving unparalleled spatial resolution (i.e., microns), and temporal information (i.e., 10^{-4} sec).

Because radar-like EM sensors are becoming so economical and so compact, they are being considering for special applications such as real time monitoring of heart motions, breathing, and vocal articulators. In this paper, it is shown that vocal fold and vocal tract wall movements can be easily measured leading to many potentially useful applications in speech signal processing, especially for cellular phones and other portable systems.

Radar-like sensors transmitting electromagnetic (EM) waves have been used to measure properties of the human vocal system during speech production (Holzrichter 1995, 1998). These sensors function by transmitting very low power (<0.3 mW) EM waves, at high frequencies, i.e., 2.3 Ghz for the experiments herein (based on measurements by D.Poland and E.T.Rosenbury, Lawrence Livermore Lab (LLNL)). A multi-cycle wave-train is directed into the neck or head toward the vocal articulators whose movements are being monitored, (see Fig. 1). The absorption coefficients at 1-4

GHz EM waves in human tissue are 5 to 10cm^{-1}, typically allowing 10 cm penetration into the body and back to the sensor (Gabriel 1996 and animal tissue measurements by L.Haddad, LLNL). EM waves reflect from all dielectric and conductivity discontinuities, associated with tissue-air or tissue-tissue interfaces, and which are in the path of the propagating EM waves. An antenna and receiver detects the reflected EM waves using a homodyne (i.e., interferometric) technique (Skolnik 1990), upon which electronic circuits average, high pass filter (e.g., AC couple), and amplify the detected signal for subsequent analysis. Sensor signals, associated with the movement of vocal folds and related air pressure induced tissue motions, occur at pitch frequencies nominally between 70 to >250 Hz. These are easily distinguishable from signals returning from stationary tissue interfaces by employing suitable frequency band-pass filters. The sensor used for the experiments herein is called a "GEMS sensor" (McEwan 1994, Burnett 1999) because it is optimized for Glottal Electromagnetic Sensing.

The signal amplitude of a homodyne sensor is proportional both to a change in distance of the targeted tissue interface from the sensor and to the reflecting locations on the tissue interface (i.e., the scattering cross-sections, see Skolnik 1990). Hence there has been an ambiguity in separating targets with large frontal areas but small position changes from those targets from those that have small frontal areas, but relatively large positional changes. In the case of the larynx region, measurement ambiguities occur if sub-glottal tracheal walls (with areas of cm^2) were moving 50-100 microns caused by air pressure changes as the vocal folds open and close. In contrast, the frontal view (area of mm^2) glottal openings is small, but their opening motion can be cm in length, perhaps

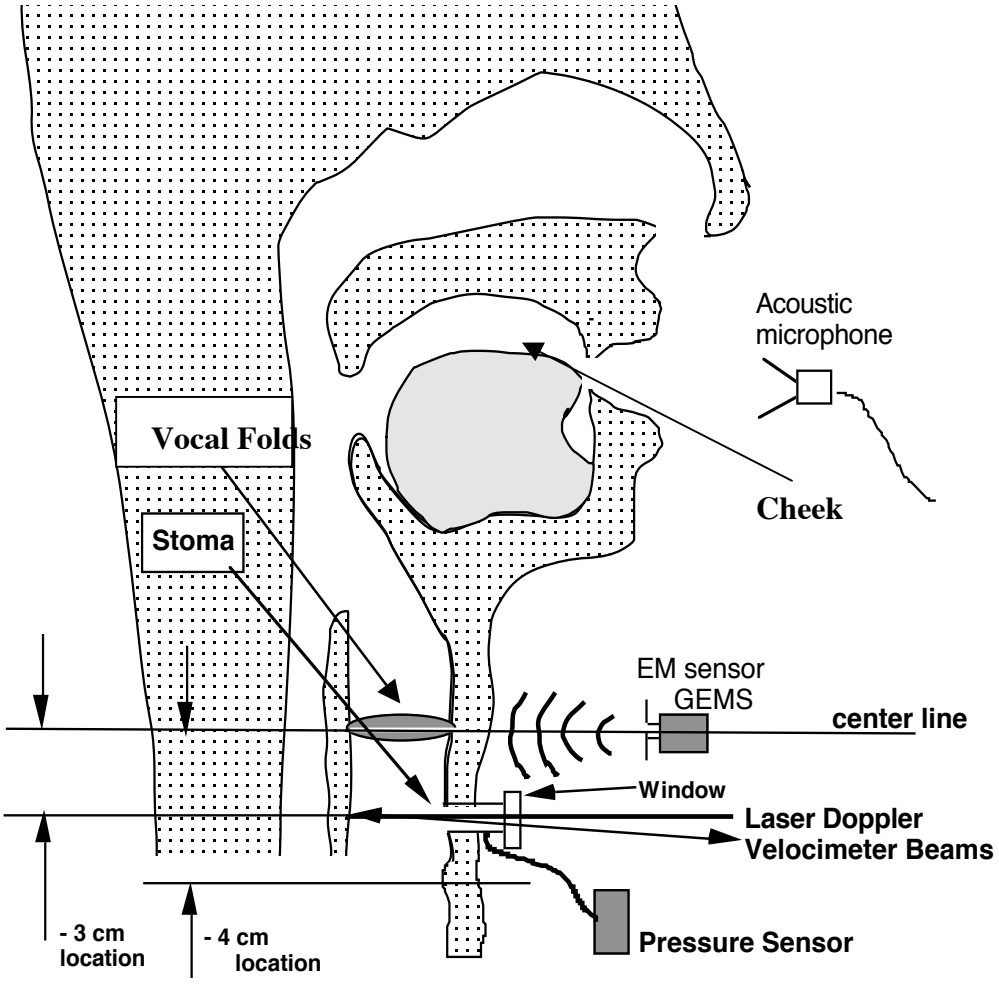

Fig. 1. Experimental layout showing position of the laser, the GEMs experimental location, pressure sensor, and acoustic microphone. EGG electrodes are not shown but are mounted normally. Three locations for EM sensor placements are noted as prominence, sub-glottal, and cheek. These locations are outside the body, against the neck skin surface.

causing a signal similar to that from the tracheal walls. In addition, when the GEMs sensor is placed in front of the larynx as voiced speech is produced, its signals resemble those from an electro-glottograph (EGG) sensor, see Fig. 2&3 (Titze 2000).

The purpose of this paper is two fold: first to identify the sources of EM sensor signals when used in the laryngeal area, and second to explain the shape of these EM sensor signals and their similarities to EGG signals. It will be shown that signals from the larynx area are due to vocal fold reflections, but signals from trachea and oral cavity walls can also be easily measured and used. Three sources of reflected EM waves from the laryngeal region are considered: 1) reflections from the changing shape of the frontal view of glottal opening, 2) reflections from tracheal-wall air-interface motions as they "balloon" due to air pressure excursions, and 3) reflections from changes in conductivity as the folds separate (i.e., EGG like). The tests used herein to reach this paper's conclusions are that the sources must lead to a reflected EM wave whose intensity versus time replicates the amplitude and then the shape of the measured signals, as seen in Figs 2 - 5.

The data and interpretations in this paper provide a great deal of information on EM wave propagation and reflection from tissues internal to the neck and head, and by extension from other articulators in and connected to the vocal tract. This understanding is important because of the potential for GEMs and other EM sensors to provide a wide range of useful, real-time information on speech type and onset, on pitch and pitch periods, and for estimates of voiced excitation functions by associating the glottal signal with air flow excitation information. These are being applied to several applications in de-noising (Ng 2000, Aliph 2002), for narrow band vocoding, and other applications.

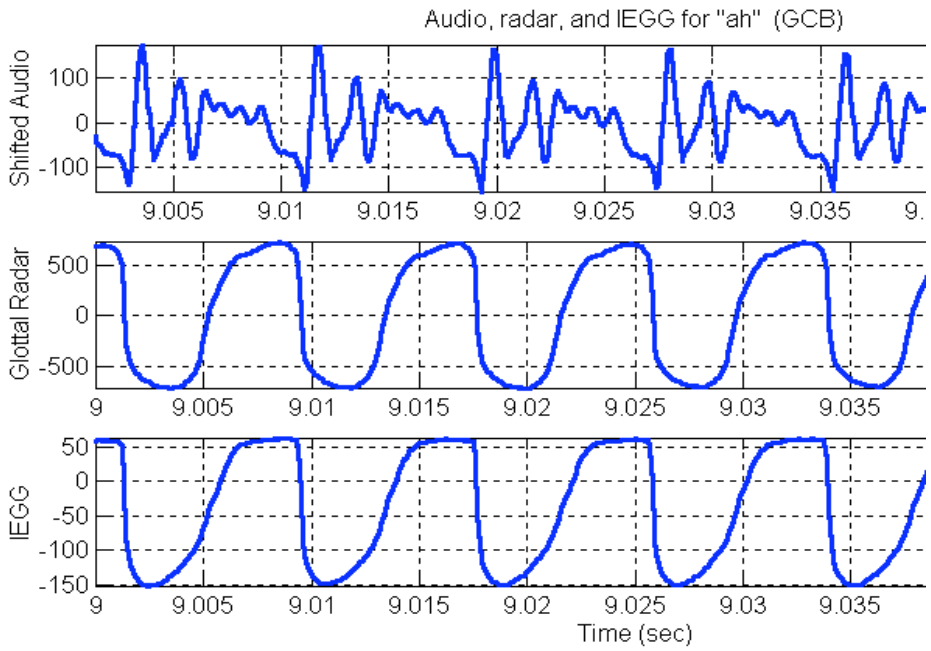

Fig. 2 Typical data showing simultaneous acoustic, GEMs sensor (i..e., "Glottal Radar trace), and Inverse EGG signals for a male speaker. Closure times are aligned. The EM sensor signals show strong similarities to those from the EGG sensors.

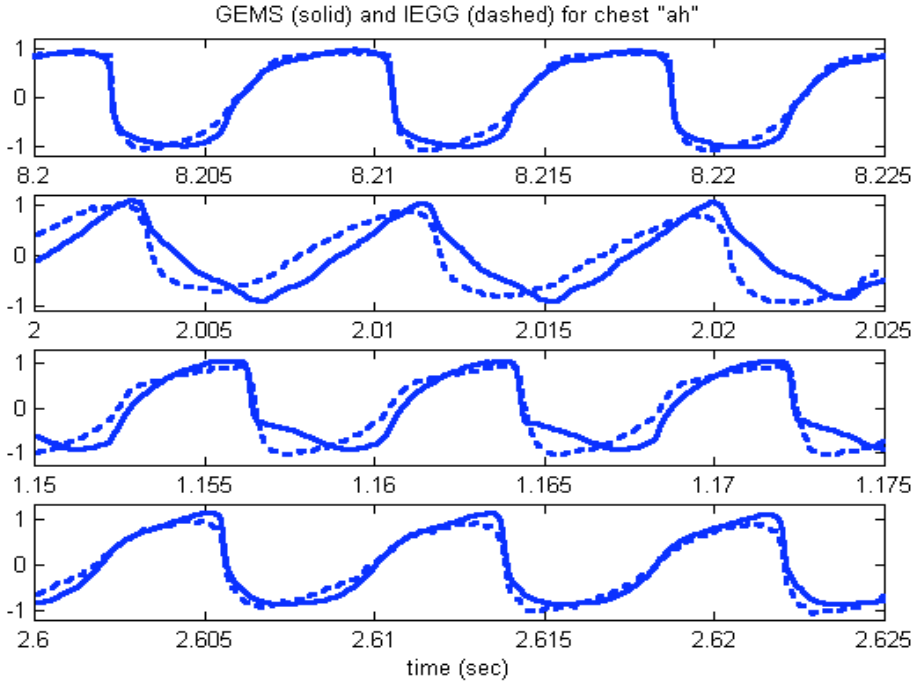

Fig. 3 Data from 4 separate male subjects comparing GEMs sensor data (solid) to EGG data (dashed). Closure times are not aligned. The idiosyncrasies between subjects and the differences between the EGG and the EM sensors are also apparent.

II. Experiments and Discussion:

The fortunate identification of an unusual subject at the Massachusetts Eye and Ear Infirmary in Boston enables subglottal pressure and posterior tracheal wall motion measurements to be made that are instrumental in resolving EM wave reflection ambiguities in the larynx region. The subject is a 58 year old female who had undergone a tracheostomy 5 years ago as a consequence of treatment for laryngeal paresis following thyroplasty with implantation of a silicone prosthesis. Her paresis was partially resolved at the time of the recording, and the subject's voice, when occluding the stoma, is only slightly breathy. The experiments described below are approved by the human studies IRBs of the Massachusetts Eye and Ear infirmary and the Lawrence Livermore National Laboratory. While the unique subject enabled unique posterior tracheal wall measurements, EM sensor and microphone signals are consistent among several dozen subjects tested to date including the subject described above.

IIA. Experimental approach:

A laser Doppler velocimeter (HLV-1000, Polytec PI, CA) is used to measure the movements of the subject's posterior tracheal wall by transmitting a laser beam through a transparent plastic cover, located over her stoma valve opening. Simultaneously the GEMs EM sensor is placed in one of two locations, depending upon the experiment. To test vocal-fold reflectivity, it is placed slightly below the thyroid-prominence on the mid-sagittal plane. To measure tracheal wall movements, it is placed below the stoma opening, about 4 cm below the prominence, also in the mid-sagittal plane. The skin is marked to provide repeatable locations for sequences of measurements. In all experiments, the EM sensor's antenna-cover (1mm plastic) is placed against the neck

tissue to prevent skin vibration. It is then gently pushed inward or outward to maximize the signal. In addition, data from an EGG sensor (Glottal Enterprises MC2-1), a subglottal pressure sensor, and an acoustic microphone (placed about 10cm in front of the lips) are also recorded, as described below. Details and calibrations are discussed in more detail elsewhere (Holzrichter 2002).

The subject habitually wears a size 6 tracheostomy tube with a Montgomery Tracheostomy Speaking Valve. For the experimental measurements, the speaking valve membrane was removed and a side opening in the valve body was drilled, to which a 6□m long side tube, 3.0 mm ID, was attached. The tube was connected to an SMI□5552-015-D pressure transducer (Silicon Microstructures Inc., Fremont, CA). To enable phonation, the open end of the modified speaking valve is covered with a transparent plastic window through which the laser was aimed at the posterior wall of the trachea. The plastic window is heated by immersion in warm water prior to placements to minimize fogging. Most of the measurements are made with the subject reclining in a supine position.

The laser instrument has been used for many physiological structure measurements in the past, including making measurements through windows and through narrow openings (Rosowski 1999). Its internal velocity calibration is quite accurate and using a given calibration of mm/second/volt, the data is converted to velocity in mm/sec, an example of which is presented in Fig. 4B. The velocity data is numerically integrated to obtain the change in position of the targeted tissue versus time, see Fig. 4C and 5D. For example in a 1 ms time interval, a velocity of 5 mm/sec leads to a movement of 5 micrometers.

The GEMs sensor's response is calibrated using an aluminum target, 1 cm wide by 2 cm high, affixed to a mechanical vibration generator and moved at 200 Hz (Bruel and Kjaer vibration generator with exciter type 4802 and head type 4817). The sensor response is found to be consistent with data obtained earlier by Burnett (1999). These procedures give a GEMs signal calibration in air of about 5 mV/micron for a target vibrating about 4 cm from the sensor.

In the case of a vibrating air/tissue interface, such as tissue vibrations from a surface located about 1 cm inside the skin (equivalent to about 5 cm of air path) the calibration is about 2.5 mV/micron. This calibration is obtained by measuring the motion of the inside surface of a subject's cheek with the laser-doppler instrument, as a subject phonates and pronounces the sound "m". This signal is then compared to the corresponding EM sensor, located 1cm away, on the cheek skin surface. Extending the cheek data to the similar geometry of the anterior air-wall interface of the trachea, results in a similar calibration of 2.5 mV/micron, or about 35 mV, for expected 15 micro-meter tissue excursions.

IIB. Experimental Discussion:

Measurements were taken on four separate occasions over the period of one and a half years. A very large amount of data was taken, with Fig. 4 illustrating one of the clearer and more complete sets of data obtained. In this experiment, five sensors are used while the subject phonates the sound "ah" as in "father". The data traces, from top to bottom in Fig. 4, are labeled as 4A through 4F, and are discussed below:

Fig. 4A shows the raw GEMs sensor output at a location just below the laryngeal prominence. This data is consistent with previous GEMs and EGG data such as that

shown in Figs 1 and 2, and elsewhere, and with the inverted EGG data (i.e., IEGG) shown below in Fig. 4E. Positive signal direction indicates increased EM wave reflectivity. As discussed in the simulations section, this signal is caused by an increased amplitude of reflection, mostly from the front edges of the glottal opening, as it opens from a completely closed condition. In addition, a reduction in the effective EM wave path and corresponding phase from the sensor to the point of reflection and back (i.e., the point of EM wave scattering moves closer to the sensor as the glottis opens).

Fig. 4B shows the velocity of the subject's posterior tracheal wall, measured 3 cm below the laryngeal prominence. At the 3 ms time, the velocity becomes negative as the glottis opens and the pressure drops. At the time of vocal fold closure, at the 4 ms time, the negative wall velocity stops increasing (in the negative direction toward the laser) and begins to slow to zero. At the 5 ms time, the posterior wall begins to move away from the laser, i.e., it "balloons", and it expands as the glottis closes and the subglottal air pressure rises. Nominal velocity noise, lasting typically 0.1 ms, indicates short periods of scattered laser light, due to scintillation.

Fig. 4C shows the numerical integration of the tracheal wall velocity trace (from figure 4B) gives the posterior tracheal wall position versus time. A positive value indicates that the posterior wall is "ballooning" in a posterior direction away from the laser sensor. The amplitude of the movement is approximately 12 microns, which is much lower than earlier estimated values used to identify the source of GEMs signals (Burnett 1999 based on intra-oral, tongue data by Svirsky 1997), but consistent with other

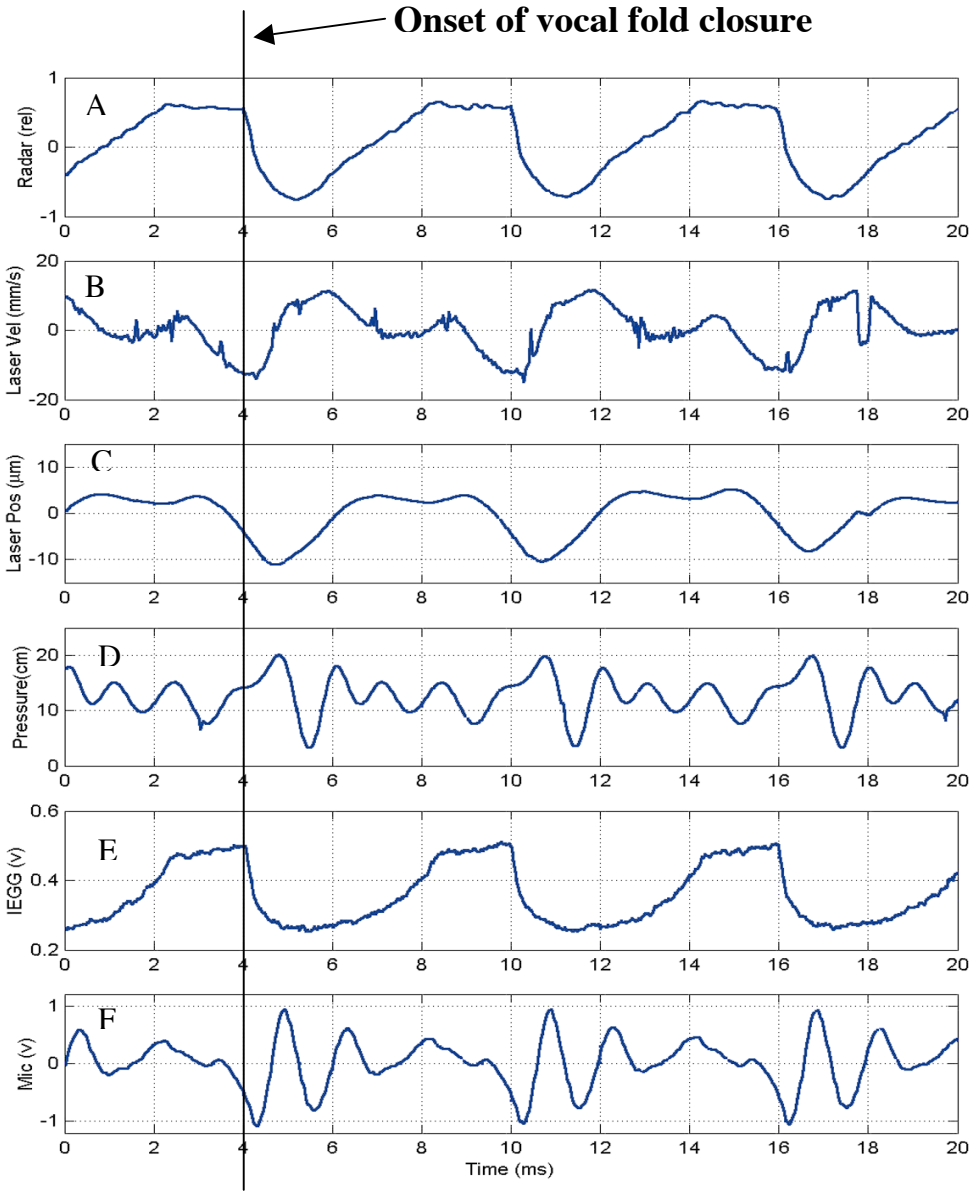

Fig. 4: Data set from subject with partial paresis but phonating "ah" very clearly. The vertical line shows the time of onset of vocal fold closure; A) GEMs sensor signal, positive value indicating increased vocal structure reflectivity; B) Laser velocity signal from posterior tracheal wall, positive signal for tissue movement away from laser (i.e., posterior direction); C) Position versus time of posterior wall, positive signal indicating "ballooning" as pressure increases; D) Subglottal pressure versus time in cm H_2O; E) inverted EGG signal; and F) Acoustic signal from microphone 10 cm in front of subject's mouth.

estimates discussed below. This tissue movement is insufficient to generate the nominal 1 V GEMS signals shown in Fig. 4A. Also, the signal amplitude and shape, corresponding to the wall movement, show little correlation with the signals shown in trace 4A) and 4E), and thus it can not be the dominant source of the GEMs signal from the laryngeal location.

Fig. 4D shows the subglottal air pressure versus time shows pressure variations of 7 cm to 20 cm of H_2O versus time, with peak pressure excursions ranging from 5 cm to 15 cm H_2O. This results in a time averaged pressure change that moves the trachea walls of about 5 cm H_2O. The signal also illustrates distinct subglottal resonances (Ishizaka 1976, Cranen 1985, and Fredberg 1978). The data from trace 4C) indicates that the tracheal wall is responding slowly and out of phase to the rapid subglottal pressure signal resonances, with about a 1.5 – 2 ms time constant. This is expected from a membrane being driven above its natural resonances (about 30-40 Hz).

Fig. 4E shows the inverted EGG signal is especially useful for timing the onset of vocal fold closure and providing information on the quality of the vocal fold contact area versus time. The EGG signal is inverted to make it consistent with GEMS data from previous papers, e.g., Figs. 1 , 2, and 11. This IEGG data follows the GEMs data closely for these and previous experiments (Titze 2000), and is used to inspect data sets to identify those with normal vocal fold closure (e.g., Fig. 4), because the subject often phonated with incomplete closure.

Fig. 4F shows the acoustic microphone signal shows the characteristic initial negative pressure signal as the folds close. The acoustic signal is advanced by 0.7 ms to correct for the slower traveling acoustic wave, from the vocal folds to the microphone

located 10cm from the lips. The characteristic peaks of the acoustic signal are used to time align the two different data sets shown in Fig. 5.

Another set of experiments (see Fig. 5D and 5E) are presented to illustrate the differences between GEMs measured lower trachea wall movement and data from the larynx location in Fig. 4. These are taken with the GEMs positioned 4 cm below the laryngeal prominence (about 1 cm below the stoma, see Fig. 1). Because of the spatial "congestion" of instruments within 1cm of the stoma, and due to some rf-interference, it is not possible to arrange for all of the instruments to be operative during the sub-glottal experiments. The data from two separate sessions are time-aligned by aligning the acoustic-peak signatures. There are three important differences between the subglottal data shown in Fig. 5D and 5E and the laryngeal prominence data of Fig. 4 . The first is that the sub-stoma GEMs signal is about 30 fold lower in amplitude, about 40 mV peak to peak in Fig. 5D, than the 1.2 V prominence signal in Fig. 5A. Using the GEMs calibration described above (2.5 mV/micron), the 40 mV signal indicates that the source is likely to be the anterior wall of the trachea, vibrating with a 10-20 micron amplitude. Secondly, the shape is consistent with a pressure-induced expansion and contraction, and shows no distinct "sharp" cutoff of the type associated with vocal fold closure. Thirdly, the timing is consistent with the laser-measured posterior-wall data of Fig. 5B, except that it precedes it in time by about 1 ms. This timing offset may be due to the quite different types of trachea-wall tissues being measured by the two instruments. Similar sub-glottal experiments have been conducted on male subjects at the Livermore Laboratory over the past years (in accord with earlier IRBs), which also show similar timing of the subglottal signal relative to the vocal fold closure signature. They also

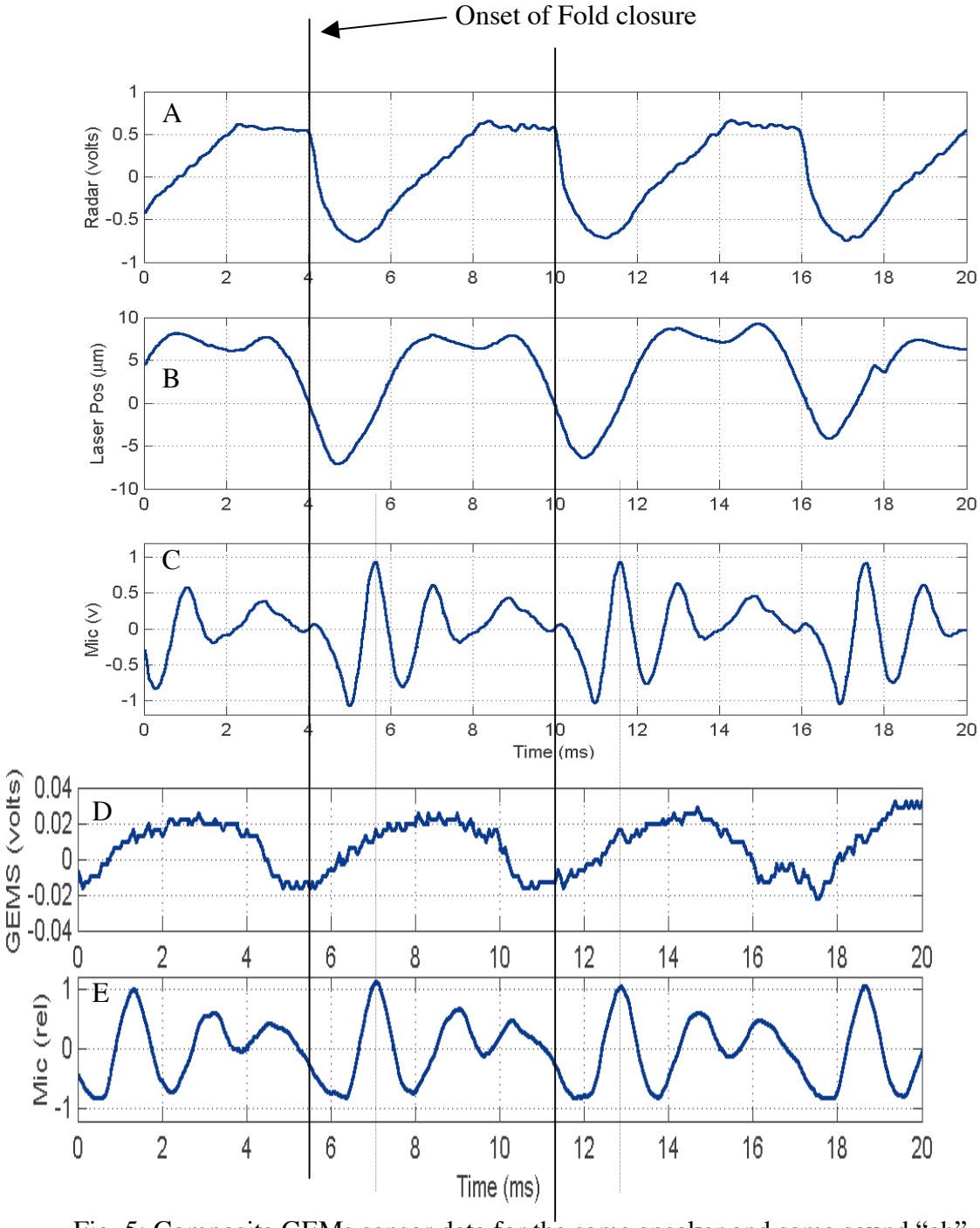

Fig. 5: Composite GEMs sensor data for the same speaker and same sound "ah" . Traces 4A,B,C are from the laryngeal area, as shown in the previous figure. Lower trace, 5D, shows GEMs signals from the subglottal area, when it was located 4 cm below the prominence. Trace 5E shows simultaneous acoustic data. These two sets of data, taken on different dates, were time aligned using the acoustic signals.

show similar relative signal amplitudes and show similar signal shape. In summary, the GEMs subglottal data are completely consistent with EM wave reflections from the anterior tracheal wall "ballooning" due to changes in subglottal pressure.

III. Numerical simulations:

Two and three dimensional numerical simulations were used to estimate the reflections of EM waves from the posterior trachea wall and from the vocal fold structures. It became clear that while 2-d calculations were useful for the investigation of EM waves interacting with cylindrical geometry, they did not enable an estimate of the effect of the vocal fold membrane and the glottal opening. For the 3-d simulation, see Fig. 6A, several "snap shots" of relatively simple geometries are chosen for the tracheal tube and the glottal membranes. The sum of reflections of the incoming EM wave, as it scatters from all the parts of the vocal fold machinery, is illustrated in Fig. 7. A simplified circular glottal hole is chosen to enable relatively simple zoning. The simplified, smooth dielectric structures, with zero conductivity, fixed glottal structures, and time integrated reflections, enable the use of a surface element code based on the "method-of-moments" (Miller 1992). For this survey, it is easier to use than using other electro-magnetic codes based on finite elements or finite differences, which are superior for more complex geometries and for time dependant behavior.

Once a few circular openings are understood, a few cases where the glottis is slot-shaped are examined (see also Fig. 6B). These slot shaped openings require more complicated zoning, especially at the closure points. The total reflections are given in figure 10, with amplitude and phase information provided in more detail elsewhere

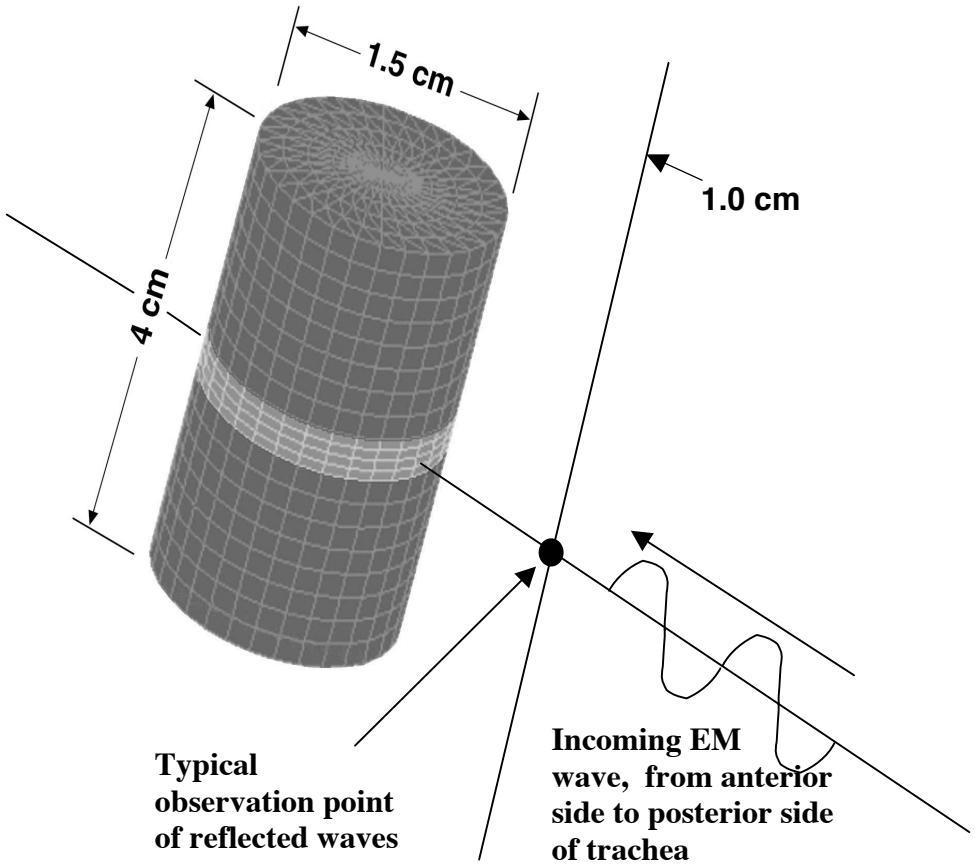

Fig. 6A: Zoning of a 4 mm thick vocal fold membrane (shown as darker ring in the center of tube) and the 1.5cm diameter trachea tube for computer simulations of EM waves propagating into the vocal fold structures. The vocal fold membrane zones are simulated as a 4mm thick dielectric disk, $\varepsilon = 25$ with no conductivity, in the mid-transverse plane of a 4cm long, "can shaped" air tube. The entire "can" structure is surrounded by an infinite dielectric material, $\varepsilon = 25$ and zero conductivity. The simulation is conducted with a time-stationary, surface element code based on the method of moments. Time behavior is calculated by computing "snap shots" of reflections as the glottal shapes are changed.

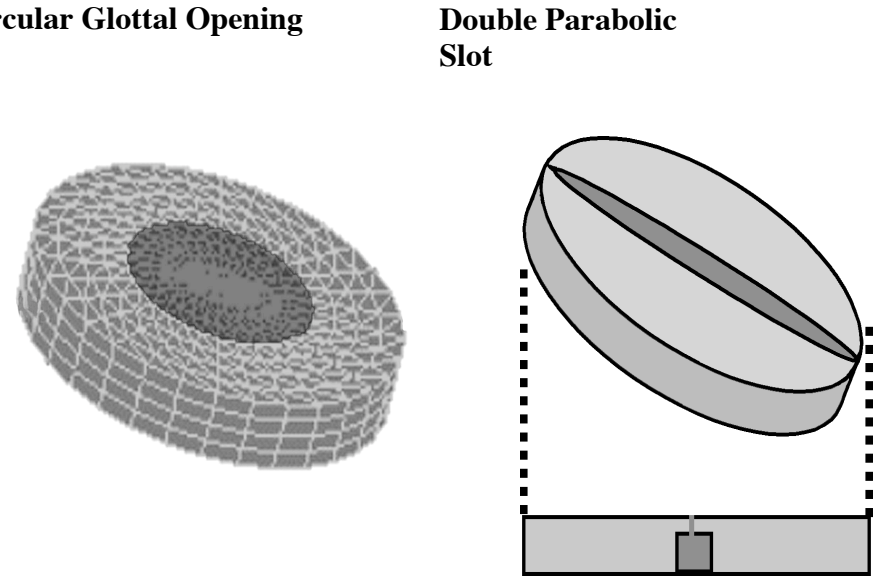

Fig. 6B: Two types of modeled glottal shapes are shown: a circular opening in the center and a double-parabolic slot, that extends from the anterior to the posterior sides of the membrane. A side view of the parabolic slot also shows a zoned side view of an incompletely opened glottis, as a phonation cycle begins. Data presented in figures in this paper were simulated using fully developed glottal openings, extending smoothly from the inferior to the superior sides of the vocal fold membrane.

(see report by Holzrichter 2003). For improved simulations, the simulated slots should be shorter in length and the zoning should capture detailed time behavior of one or more observed modes of vocal fold opening and closing, as the margins open from the inferior side, ultimately to the superior margins, and then close.

IIIA. 3-D simulation approach:

A "Method of Moments" code, called EIGER (Sharpe 1997), was chosen for the survey calculations because of its availability, its reliability, and the presence of computers able to deal with the relatively large number of computations needed for applications of this procedure (Miller 1992). This code was used to analyze the propagation of an EM wave internal to an infinite block of neck tissue. The electrical properties of the layers of muscle and cartilage were approximated by an average dielectric constant, $\varepsilon_{average} = 25$ and zero conductivity (no loss in the tissues). Fig. 6 illustrates the shape of the modeled larynx region, coded using 1500 nodes. (In some cases, such as slot calculations more extensive zoning is used). For the method of moments, the trachea air tube requires proper boundary conditions which are satisfied by using "end caps" to keep the simulation volume finite in extent, and by extending the surrounding dielectric material out to infinity. Contributions to the E field at the point of measurement, from EM wavelets scattering from the caps, were tested to be sure that they did not contribute significant signals at the measuring point. In addition, the zoning of the membrane, with its various holes and slots, was varied to be certain that the amplitudes and phases of the simulated EM wave reflections were not caused by numerical artifacts.

A 2.3 GHz, EM plane-wave-train, y-polarized (in the plane of the membrane), with an amplitude of 75 V/M was "launched" from one side of the simulation (i.e., from within the dielectric coming from the right side in Fig. 6). The wavelength in the dielectric is λ = 2.6 cm. The EM wave propagated in the dielectric medium to the front surface of the 1.5 cm diameter trachea (i.e., from the lower right of Fig. 6 into the trachea section, which is 1.5 cm in diameter by 4 cm long. The EM wave then propagated (i.e., evanesced) around the tube. Normally an EM wave cannot enter an air cavity with a dimension smaller than the wavelength in the surrounding high dielectric-constant media, Fig. 6A. However, as the vocal folds are adducted into the trachea tube, the tissue membrane "wave-guides" the EM wave across the air tube, causing the EM wave to follow the high dielectric-constant material. At the same time the EM wave was propagating forward into the tube, it was also being reflected by the front and back walls of the tube, and by glottal holes or slots in the membrane. Fig 7 shows 4 examples of the scattered component of the EM wave, where it is being forward-scattered across the membrane and back-scattered to the GEMs sensor. The numerical simulations also enable an estimate of the reflection of an incoming EM wave from a nominal glottal opening back to the sensor to be about 0.5 %.

The Eiger code's numerical diagnostics enable simulations of both amplitude and phase of the scattered incoming EM wave at a desired observing point. This point is located 1cm in front of the anterior wall of the trachea in the mid-transverse plane (on the right side of the images, on the side of the incoming EM wave). This is commonly the

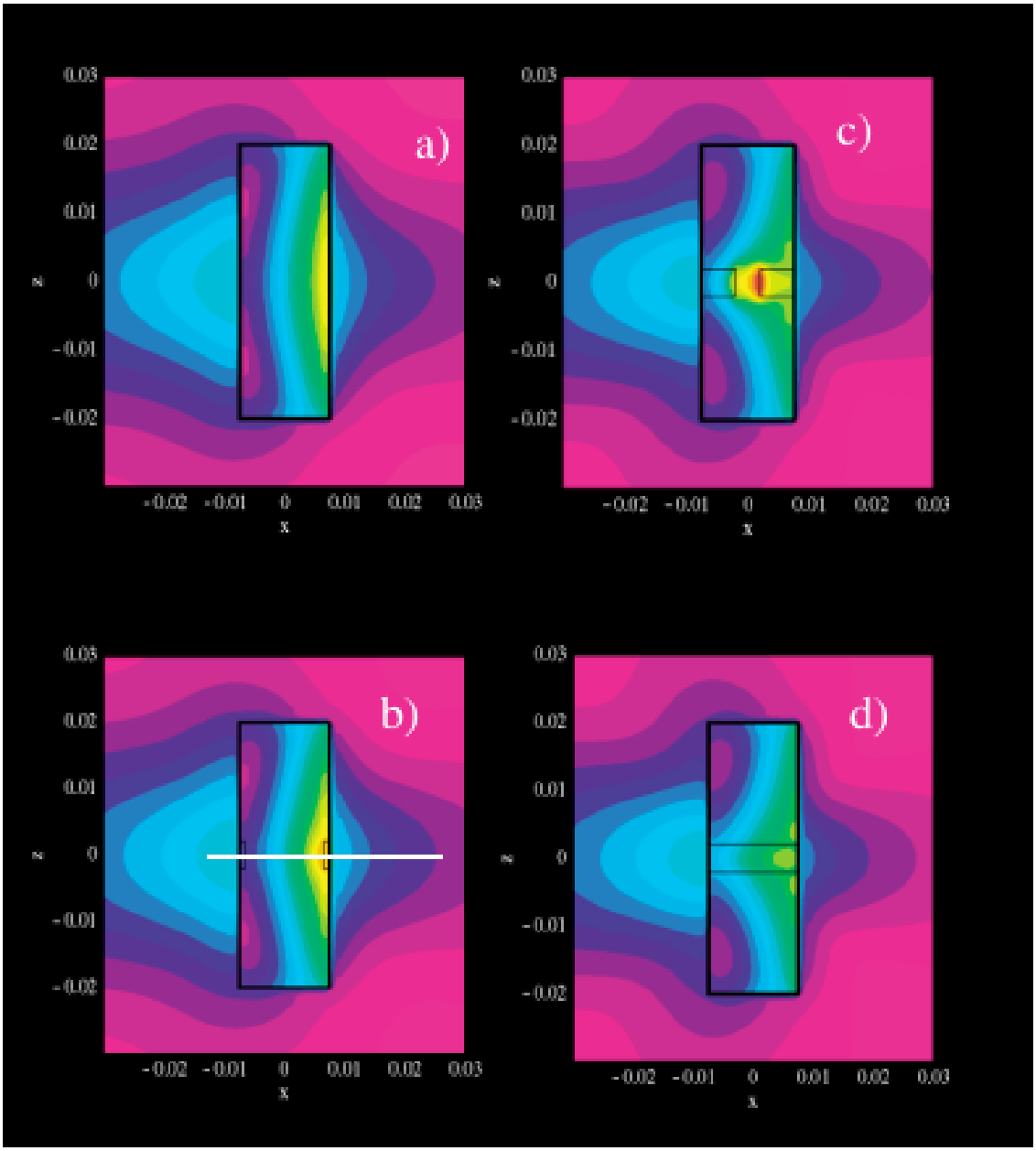

Fig. 7 : Four examples of simulations showing the amplitudes of reflection of incoming EM waves (from right to left) onto 4 different tracheal tube sections, each with a different diameter circular glottis. The four cases, a-d, have corresponding glottal diameters of 1.5, 1.2, 0.4, and 0.0 cm diameter respectively. In cases b,c, and d the outline of the cross section of vocal fold membrane is in black. The white line in insert b shows the axis of "line-out" data showing scattered field-amplitude and field-phase versus position (shown in following figures). From "line-out" data the color code follows: red at the edge of the hole in c) corresponds to a field strength of 180 V/M, yellow 150V/M, green 130V/M, light blue 75 V/M, and the dark-purple color, on the right of the tubes, 40 V/M.

place in the experimental configuration where the EM sensor receiver antenna would be located on the skin on the laryngeal prominence. While this diagnostic point is useful for survey work, a real antenna extends several cm laterally, measuring many locations where the signal can change significantly from the on-axis value. The numerical data was used primarily in the differential mode, where the input wave is held constant and changes in vocal fold configurations cause phase and/or amplitude differences in the scattered EM waves, which are compared, one to another.

IIIB. Simulation results:

Many numerical and empirical experiments show that the sensitivity of the EM sensor signal, corresponding to positional and directional variations of $\pm$ 1 cm and $\pm 10^0$ respectively, is small. This phenomenon occurs because the EM wave is about 2-3 cm wide inside the neck tissue, allowing it to be easily "captured" by the vocal fold membrane and guided across the trachea. The images shown in Fig. 7 also show color contours of scattering EM wave amplitudes versus distance in the mid-sagittal plane. Fig 7A illustrates EM scattering from the tube with completely retracted folds, showing how the EM wave weakly penetrates the anterior tube surface, but strongly reflects backward from it. In the intermediate cases, Figs 7B and 7C, red and yellow colors, indicate that the EM wave is "carried into" the tracheal tube along the membrane, and scatter from the posterior edges of the glottal opening. Figure 7D illustrates how the EM wave is carried across the tube by the 4 mm thick dielectric membrane with a low level of reflection.

Figs. 8 and 9 show the amplitudes and phases of the reflected EM wave at a continuum of measurement locations along a path from the trachea center to beyond the measurement point of the hypothetical experiments (see white line in Fig. 7B illustrating the path). This point is located 1 cm in front of the tube (at the 0.0175 M location) and is designated by an arrow on the graphs. The glottal model, used to simulate the results reported here, consists of a simple expanding and contracting circular hole in a 4mm thick membrane (e_{Ave} = 25), starting with a solid membrane with area ratio = 0. The area ratio, AR, is given by the ratio of glottal-area/trachea-area = area ratio. Next a hole with relative area ratio AR=1/4 is simulated, and finally a hole with an area ratio AR= 1/2 . The condition noted as AR=1 is the empty tube. The circular glottis model also enables simulations of empty-trachea wall reflections, which occur when the glottal diameter becomes equal to the trachea diameter (i.e., 1.5 cm in these examples). In the model, as the glottal opening is increased in the simulation, the glottal membrane is thickened to represent the volume of tissue that moves to the sides as the glottal space opens. This approach conserves tissue molecules in the path of the EM wave.

Fig. 8 shows that the reflected field amplitude, on axis, ranges from 42 V/M for closed folds, 45 V/M with AR = 0.25, 47 V/M with AR=0.5, and 48 V/M for an open tube trachea. For this example, at the measurement location, the reflection amplitudes change by 12-15% from the closed glottal condition to the open condition. The background amplitude of 42V/M is the scattered amplitude from all the modeled reflecting structures, such as the tracheal walls, the end caps, and the membrane-wall interfaces. Since these are constant, they don't show up in high-pass filtered signals.

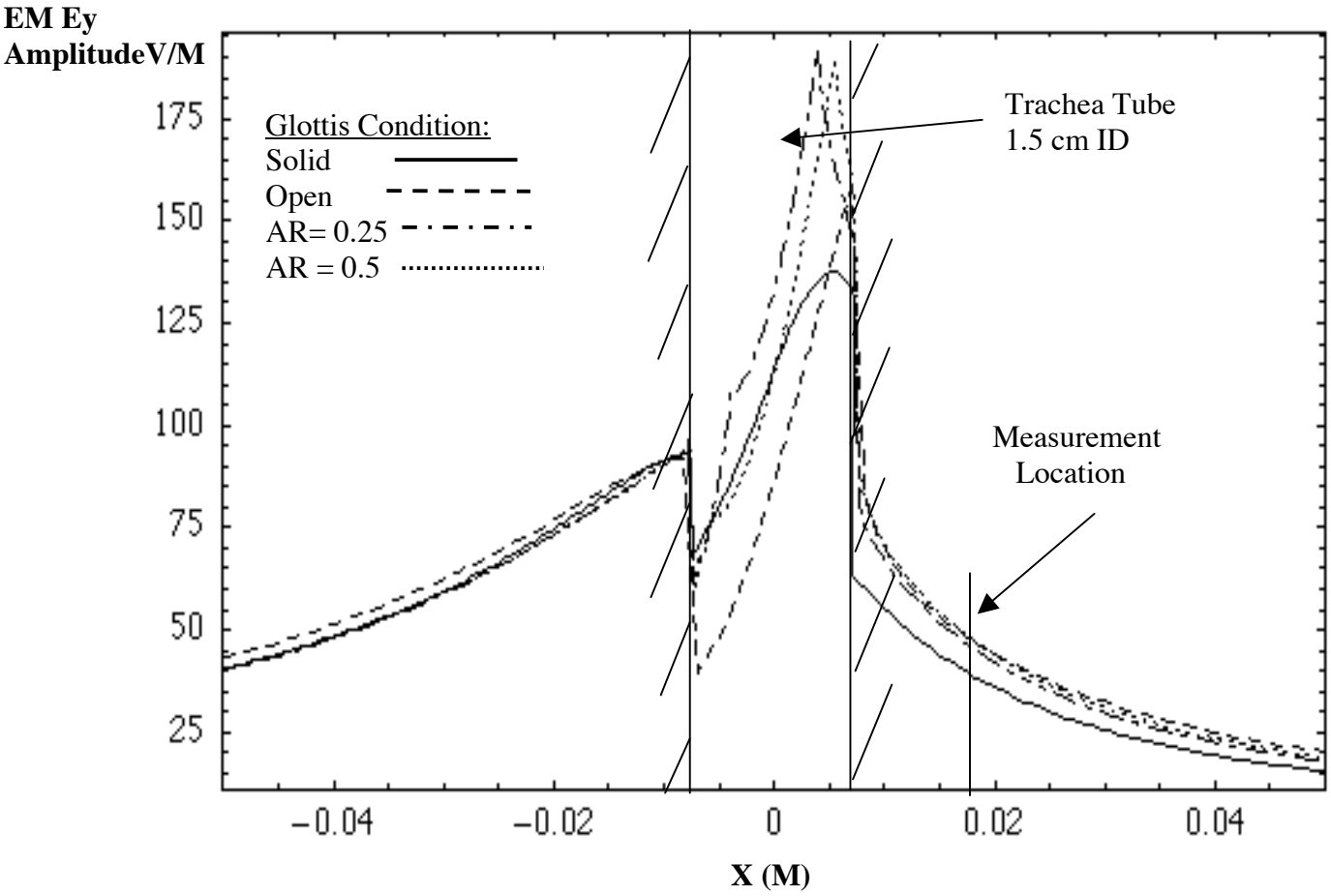

Fig. 8: Amplitude of the forward and backward scattered EM waves, both inside the trachea tube and outside the tube, for 4 different circular glottal openings. These 4 curves follow the scattered EM field amplitude along the mid-sagittal center line of the calculations. They give the data "line-out" along the white line in the previous figure (in Fig. 7b). At the noted "measurement location" , the intensity of the reflected EM wave depends upon the structure of the reflections in the glottal membrane, as well as surrounding structural elements. The solid line is for a closed glottis in a 4mm thick membrane, the dot-dashed line is for a 0.75 cm hole (area ratio = 0.25), and the dotted line is for a hole diameter of 1.06 cm (area ratio = 0.5), dashed line is completely open. Signals are interpreted as differences between changing and stationary reflections.

Fig. 9 shows the reflected field phase versus distance, due to glottal scattering condition changes. The measurement location is noted. The convention in the "Eiger " code is that the phase is increasingly negative as the distance increases from the effective source of reflection to the measuring point. For example at the noted measurement location, the phase is increasingly negative as the glottis closes and the wave continues to the far side of the trachea tube. This means that a fraction of the EM wave is carried deeper into the trachea tube as the glottis closes, thus increasing the path of the wave. The phase changes for the four glottal conditions shown are -36^0, -27^0, -25^0, and -8^0 for the empty tube, giving a maximum phase change of 28^0, or a 33% phase change within a 90^0 quadrant of phase. As is summarized in Fig. 10, the phase change from a slot opening behaves differently than that of an expanding circle because the front edge of the slot scatters most of the energy and defines the reflection location, thus fixing the phase. The phase changes little as the slot widens because waves reflected from the sides do not come back to the measuring point.

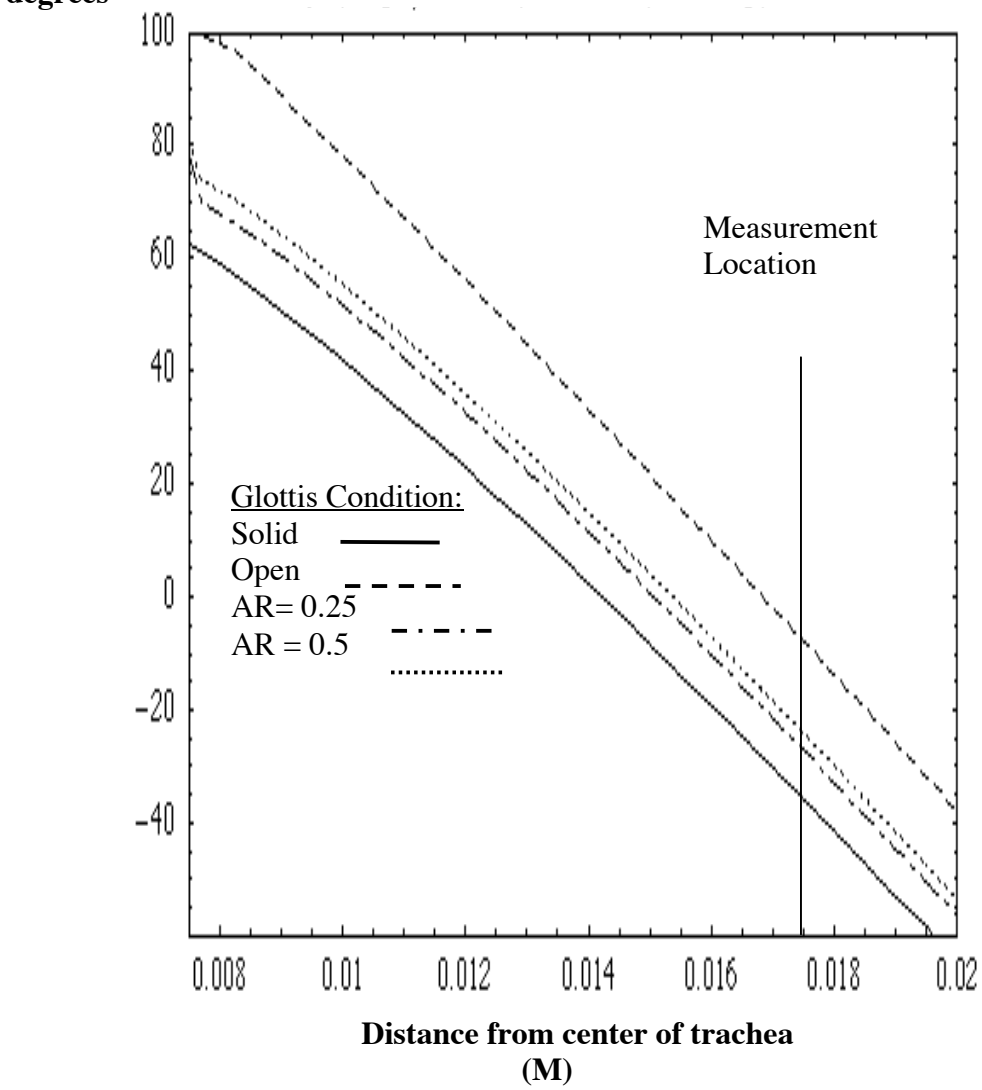

Fig. 9: Phase of the forward and backward scattered EM waves, in degrees, versus distance and versus glottal condition, measured from the trachea tube center to the "measuring point". The left ordinate line is positioned at the anterior trachea wall. The phase increases (becomes more negative) when the glottis closes and "draws" the EM wave further inside the trachea tube.

IIIC. Comparison of sensor signals to glottal area

The GEMs sensor signal is estimated by taking into account the output of the signal from a homodyne receiver detector (Skolnik 1990) as shown in equation (1):

$$S_{mixer} = \text{const.} * A * \cos \varphi \qquad (1)$$

A is the received signal amplitude and φ is the phase change as the tissues moves. For maximum sensitivity, this phase is usually referenced to the tissue rest position, at which point the phase is adjusted to be $\pi/2$ (Burnett 1999). Hence the "AC-coupled" sensor signal, which is function of both the phase change and the amplitude of reflection, is estimated by comparing the signal from a final configuration to that of an initial configuration.

Fig. 10 shows relative, on axis sensor signal levels for two types of glottal openings, a parabolic slot and a circular shape (see Fig. 6B for the geometries). The data is plotted as signal versus area ratio (i.e., where the area ratio, AR, is the area of glottal opening divided by tracheal tube area). This presentation is intended to show to what degree glottal air flow, proportional to glottal area (Stevens 2000, p. 65), can be related to the EM sensor signal. This approximation, that the vocal tract excitation source is proportional to glottal area, is commonly used for signal processing applications (Ng 2000). The signals measured by the EM sensor are the modulated signals from the moving tissues, which pass the filter bands, and which are superimposed on top of a non-modulated background (e.g., 42 V/M in the example above). The circular glottal opening begins as a small hole in the middle of the

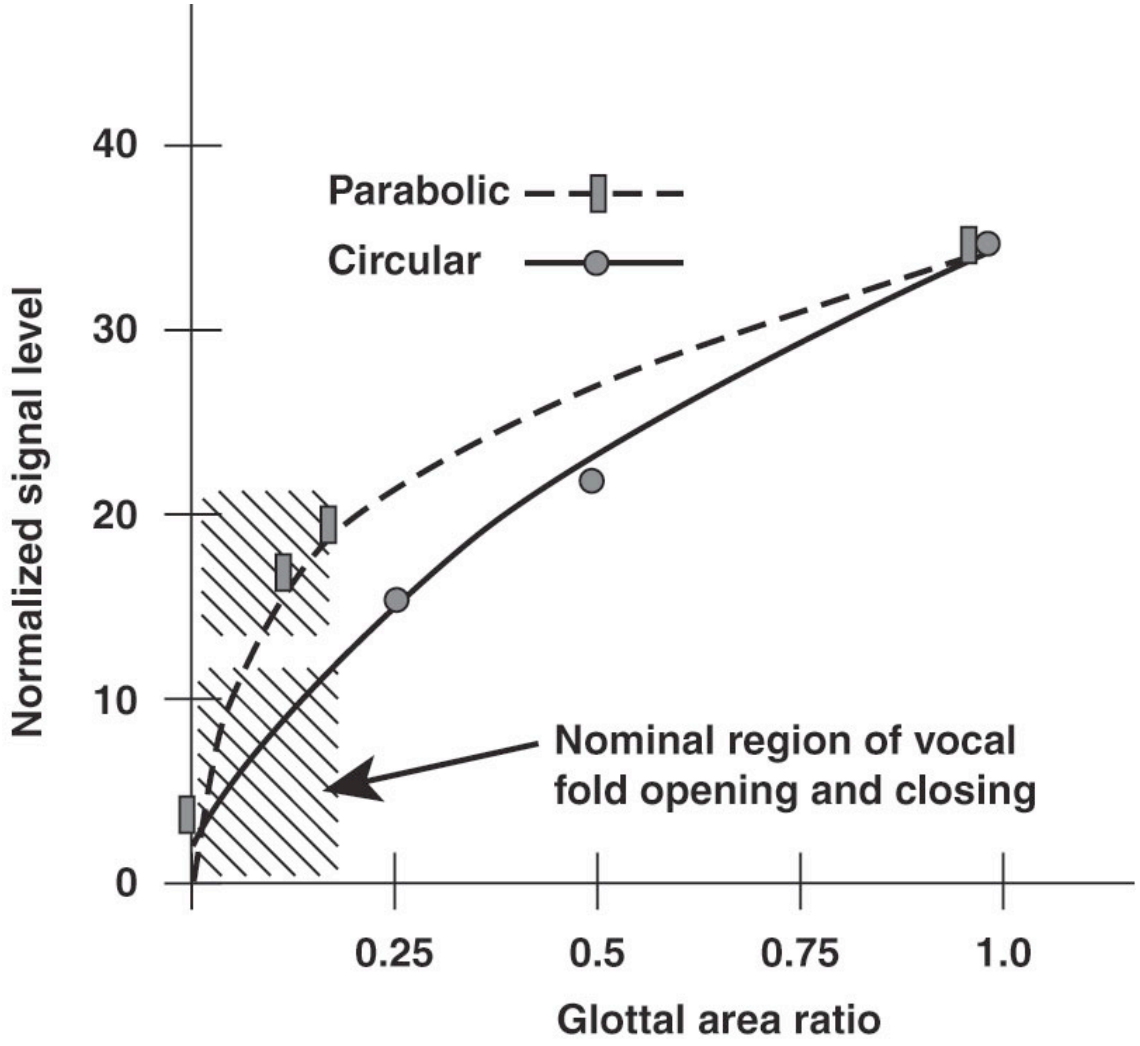

Fig. 10: Smoothed and normalized EM sensor signal data from simulations of reflected EM waves from circular glottal openings and from parabolic openings versus AR. AR is the area ratio showing the relative area of the glottal opening to the tracheal cross sectional area. The data takes into account both the field amplitude and phase change as the glottis opens and closes. The zero signal point, corresponds to the reflected signal of 42 V/M, from all reflections in the simulation. Similar, stationary reflections occur when an EM sensor is placed against the neck or other feature of a user.

membrane, scattering a small amount of the incoming EM wave. As it increases in diameter the reflection increases, and at AR = 0.2, its signal modulation is (52-42)/52 = 0.21.

For a slot, the opening extends across the diameter of the tube (see Fig. 6B) from the anterior to the posterior walls of the trachea, and its area is related to the width of the parabolic edges. Simulations show that upon the slot opening, the EM wave reflects strongly from the anterior point of the slot shaped opening, showing an amplitude and phase jump. As the slot widens further, Fig. 10 shows that the reflectivity increases more slowly than for the circular opening. At an AR value of 0.2, the parabolic slot modulates the reflected EM sensor by over 30% (e.g., signal modulation is (62 – 42) / 62 = 0.33). In the parabolic case for the small glottal area regime (e.g, 0< AR < 0.1) a signal proportional to glottal area, and thus to glottal air flow (Stevens 2000, p. 65), could be constructed from data in Fig. 10. However more detailed calculations are needed.

Actual glottal shapes resemble a combination of the two model structures discussed here, but with several additional complexities. For most persons, the glottis is shorter in length and does not extend across the tracheal tube. Upon opening the slot evolves to an elliptical shape (e.g., see Fig. 11). In addition, the glottis opens as the inferior margins of the vocal folds slowly open, and then close relatively rapidly, mucosal motions take place that also reflect EM waves, incomplete posterior closure is common, and many other phenomena occur (Rothenburg 1981, Titze 1984). These types of varying signal sources will be discussed below in conjunction with Fig. 11.

IV. Discussion of data and simulations:

The simulations in Figs 8-10 enable a calibration of the GEMs sensor's signal as a function of the modeled changes in phase and amplitude of the reflecting EM wave from reflecting tissues. First, it is estimated that for each unit of EM wave amplitude transmitted toward the glottis from the anterior skin surface of the neck, about 0.005 (i.e., 0.5%) is reflected back to the sensor's antenna, or in other words the total received power is about –42db. Typically 1 to 10% of this reflected signal is modulated by moving tissue interfaces. Next, in order to compare calculations to measured signals, on a relative basis, consider the hypothetical sensor, located just below the laryngeal prominence at the "measuring point" (see Fig. 6A). For a nominal glottal area ratio change, e.g. from 0.0 to 0.20 in Fig. 10, the EM sensor signal is about 1 Volt (see Fig. 4A). This signal is caused mostly by a phase change of about 14^0 , causing a 25% signal change (as shown in Equation 1), but it is also due a small reflected amplitude change of about 8% (circular model, Figs 8&9). For comparison to calculated signals from tracheal wall tissues whose signal is mostly due to small phase changes, the GEMs sensor phase sensitivity is estimated by equation (2):

$$C_{glottal} = 1 \text{ V}/14 \text{ deg phase change} = 0.07 \text{ V/deg} \qquad (2)$$

IV-A: Tracheal wall results:

EM waves reflected from both the posterior wall and the anterior wall are considered. These are important for applications because they represent other vocal tract wall movements, including the pharynx, oral cavity, and sinuses, that can be measured and used for signal processing applications. The direct measurement of the sub-glottal, posterior tracheal wall tissue with the laser velocimeter shows 12 micron movements

(see Fig. 4). The opposing anterior tracheal wall movement was not initially considered (Burnett 1999) because of its expected stiffness. However, a direct examination of a human trachea, excised from a cadaver, showed that the stiff "C" shaped, trachea cartilage structural elements occupy about 30% of the anterior tracheal wall area. The remaining 70% of the wall area, between the cartilage elements, was found to be similar to other vocal tract wall tissue, and would move similarly to other wall tissues, as the subglottal pressure oscillates.

1. Anterior trachea wall

Using the calculations shown in Fig. 9, for an open tube, an estimate of the tracheal wall phase delays due to small wall movements can be made by moving the numerical observation point by a small distance, see (Holzrichter 2002) for details. The anterior tracheal wall calibration is shown in equation (3):

$$C_{anterior} = (0.44^0/10\,\mu m)*(1V/14^0) = 0.030\ V/\ 10\ \mu m \qquad (3)$$

The GEMs signal from 4 cm below the prominence, Fig. 5D, is 35 mV, which upon using the calibration of 30mV/10 microns (Eqn. 3), yields a simulated movement of 12 microns.

Earlier, the calibration of a signal for a similar geometry was discussed, that of the inner check wall, which was measured using a laser velocimeter. This calibration gave 2.5 mV/ micron of inner cheek surface motion for a 5 cm H_2O intra-oral pressure change. Using the measured 30-40mV GEMS sub-glottal signal (see Fig. 5D & 5E), and the cheek-based calibration of 2.5 mV/micron, an anterior trachea inner-wall surface movement of 14 microns was measured. This agreement between the simulated anterior wall signal of 12 microns and the similar cheek motion of 14 microns is sufficiently

close that it is concluded that a GEMs-like sensor, when placed at the –4 cm subglottal location as in Fig. 5D and 5E, measures movement of the anterior tracheal wall.

2. Posterior tracheal wall

The posterior wall induced changes in the GEMs signal can be simulated in a fashion similar to those just used to estimate the anterior wall signal, and can be compared to directly measured signals as shown in Fig. 4C. . Equation (4) shows its calibration:

$$C_{posterior} = 0.018 \text{ deg/10micron} * 1V / 14 \text{ deg.} = 0.0013 \text{ V / 10 micron} \qquad (4)$$

Using the laser-doppler system, 13 microns of posterior wall motion were measured (Fig. 4C). Using the calibration just above in Equation 4, the GEMs signal from posterior tracheal wall motion would be about 1.7 mV. This signal is substantially smaller than the GEMs measured signals from both the laryngeal and sub-glottal region, indicating that the posterior trachea wall motion does not contribute signifantly to the GEMs signals in Figs. 1-5.

Because of the small measured motions of the posterior tracheal wall, which conflicts with some earlier estimates of air pressure induced intra-oral tissue measurements that were extended to tracheal tissue responses (Burnett 1999), some additional estimates of the tracheal wall motions using mechanics formulae are made. A standard thick-walled pipe formula (e.g., Roark 1975) was used with an inner radius a = 0.75 cm, and an approximate outer radius b = 2.5 cm. Then the change in the inner radius due to pressurization is given in equation (5):

$$\Delta a = (\Delta \text{ pressure}) * (a / E) * [(b^2 + a^2)/(b^2-a^2) + \nu] \qquad (5)$$

The radius a is 0.75cm, the Δ pressure is 5 cm H_2O, Young's modulus E for smooth muscle is estimated to be about 190 psi (Fung 1993), and Poisson's ratio is $\sigma = 0.2$. This calculation leads to a cyclic wall deflection of about 5 microns, due to air pressure excursions of about 5 cm H_2O. A second calculation, modeling the posterior trachea wall as a compressible plate, using a unit strain formula of $\varepsilon = \Delta / E$, yields a movement of approximately 10 micrometers per 2 cm thick wall for a 5 cm H_2O pressure. In addition, a tissue compliance (i.e., spring-like "K" values) can be estimated from the posterior tracheal wall measurements in Fig. 4 (i.e., 13 microns per 5 cm H_2O pressure change) which gives a value of about $30x10^5$ dynes/cm. This measured value is however substantially stiffer than exterior side-neck tissue compliance data from Ishizaka, $5x10^5$ dynes/cm (1975). This Ishizaka neck data is, in turn, about 10-fold stiffer than tongue tissue data presented by Svirsky, $0.5x10^5$ dynes/cm (1997), upon which early tracheal wall motions were estimated. In summary, the static stiffness estimates above, the computer simulations, Ishizaka's data, and the experiments in this paper show that the posterior tracheal wall movement is approximately 5-15 microns, which is difficult to detect using presently configured GEMs like sensors.

V: Comparison of EM sensor signals to laryngograph images

It is useful to compare the superior view of the vocal fold surface tissues (Burnett 1999, Burnett and Leonard 1999) to the corresponding EM sensor signals. See Fig. 11 for an example of data. The descriptions are aided by vocal fold cycles as illustrated in Stevens (2000). It is interesting to note that at location "A" the EM signal is rising

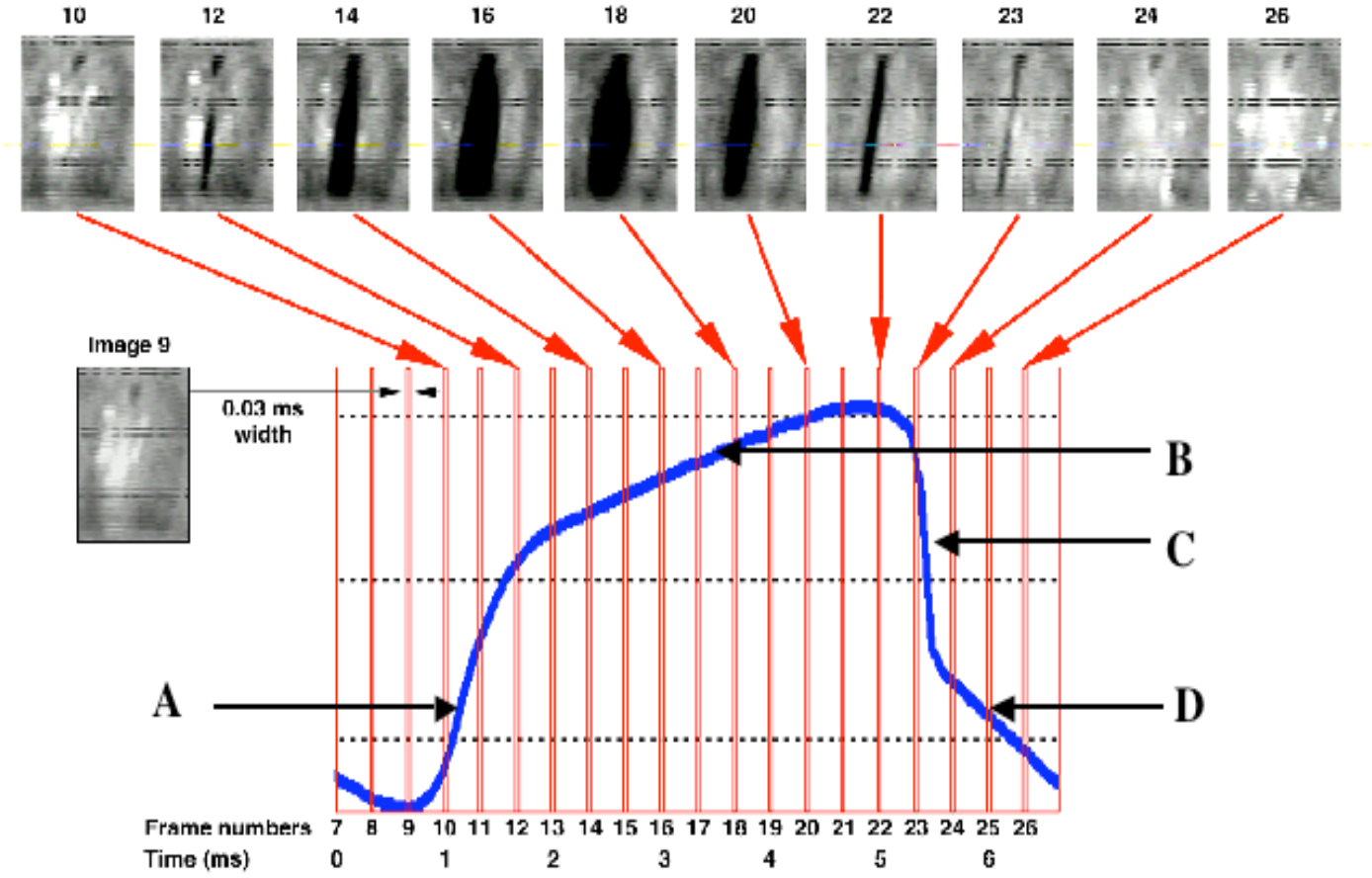

Fig. 11: Typical GEMs signal from a male subject with corresponding vocal fold images. Location A shows rapid sensor signal increase as the inferior edges of the fold open increasing the EM wave reflectivity more rapidly than linear from a parabolic-like slot, location B shows signal increasing nominally with area increase as in an elliptical-circular shape, location C shows rapid signal drop, probably due to strong anterior fold contact, and location D shows continued sensor signal drop, after the visual image shows fold-closure, indicating continued vocal fold tissue rearrangement in preparation for the next cycle.

relatively rapidly as the inferior margins of the vocal folds "peel" open from the inferior side of the membrane (see Fig. 6B slot picture for a model example). The increase in vertical slot opening causes a more rapid than linear increase in signal, estimated to be at least proportional to the square of the slot thickness (Born 1965), because EM waves reflect poorly from structures that are small compared to their wavelength and reflect strongly as the dimensions approach 1/4 wave-length and larger. In addition, numerical simulations show that tissue conductivity is not needed to generate these rapidly changing signals, although it should be considered in future calculations. The signal in region B increases less rapidly than that in region A because the slot is already formed, and the targeted opening has become more circle-like, see Fig. 10. However, the images also show that closing begins at the 4 ms time marker, yet the EM sensor signal continues to rise. This remains puzzling, except that the EM sensor is measuring the changing configuration of all of the tissues in the laryngeal region, including the vocal fold mucosa, processes, and ligaments. (However, in the case of the subject whose EM sensor data is shown in Fig. 4A, this pattern is not seen). The rapid closure signal at location C is associated with vocal-fold contact, most likely with initial contact in the anterior region, where the contact location and extent most strongly influences the phase and amplitude of the EM wave reflection. The rapid closure signal in location C is often used for signal processing applications, because it provides accurate timing and because it empirically behaves well as an excitation function, generating most of the spectral energy in the voiced speech signal. The signal may be proportional to air-flow cut-off as discussed in Section III above, and thus is an adequate representation of a voiced excitation. The more slowly relaxing signal in region D, indicates that for this subject

(also see subject 3 in Fig. 3), complete closure of the folds and their rearrangement for the next cycle takes more time than normal. This example illustrates the sensitivity of radar-like EM sensors to many idiosyncratic factors of a subject's phonation, which can be used for speaker verification parameters (Gable 2000). In addition, these EM sensors easily measure the state of folds during falsetto, breathy speech, and many other conditions where vocal fold contact does not take place. The EM sensor is sensitive to the size, shape, dielectric constant, modes of motion, and the time dependant arrangements of the vocal folds and surrounding tissues, and not strongly sensitive to vocal fold contact conductivity.

VI. Conclusion:

Based upon the experiments and simulations described herein, homodyne EM sensors can measure real-time motions of the vocal folds and vocal tract walls, such as those of the trachea, cheek, and other locations. However the signals from these two general types of tissues are very different in both shape and amplitude. EM sensor signals, measured just below the laryngeal prominence, are due to reflections from an EM wave guided by the vocal fold membrane, which samples the changing configurations of the vocal folds. In contrast, EM sensor measurements of air pressure induced sub-glottal, tracheal wall movements, and intra-oral cheek walls, are shown experimentally and analytically to be associated with a nominal air-pressure induced 10-20 micrometer movements of the anterior tracheal walls.

Data from the EM sensors of the type illustrated here are usually combined with corresponding acoustic data for speech processing applications. For example, they

enable robust, extremely accurate onset of voiced speech, pitch detection (< 1 Hz accuracy), and an estimate of a voiced excitation function by using the spectral information associated with rapid fold closure (e.g., signal segment C in Fig. 11). They enable acoustic signal de-noising of at least -20 dB reduction, depending on type of noise (Ng 2000, Aliph 2002). They also appear to enable narrow bandwidth speech compression (200-400 Hz bandwidth), and experiments indicate that speaker verification is possible with error rates less than 1:1000 (Gable 2000). These relatively new EM sensors are becoming increasingly economical, use very little power, and can be built into a variety of instruments and commercial devices.

Acknowledgements:

We would like to thank Dr. Greg Burnett for many early suggestions and measurements. We would also like to thank Drs. Rebecca Leonard, Wayne Lea, and Professors Ingo Titze and Brad Story for their continued interest and support. The work at LLNL was supported by the US Department of Energy, DARPA, and the National Science Foundation. Work at the Massachusetts Eye and Ear Infirmary was supported by NIH/NICDC.

REFERENCES:

Aliph (2000), applications of EM sensor/acoustic speech technologies. Aliphcom Incorporated, San Francisco, CA, see examples at www.aliphcom/sound (best accessed using Microsoft Explorer)

Born, M. and Wolf, E. (1965) "Principles of Optics" Pergamon Press, Oxford

Burnett, G.C. and Leonard, R. (1999) "Use of Kodak Ektapro High-Speed DigitalCameras in Laryngoscopy" Phonoscope (2) 1, 33 (1999) Singular Publishing Group

Burnett, G.C.,(1999) "The Physiological Basis of Glottal Electromagnetic Micropower Sensors (GEMS) and Their Use in Defining an Excitation Function for the Human Vocal Tract" Thesis UC Davis, Jan. 15th, 1999, available through ProQuest Digital Dissertations, document number 9925723.

Cranen, B. and Boves, L. (1985) "Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models for speech production" , J.Acoust. Soc. Am. 77 () 1543 (1985)

Flanagan, J. L.,(1965) "*Speech Analysis, Synthesis, and Perception* ," Academic Press, Inc., New York. See pp. 40-41. Also 2nd edition 1972.

Fredberg, J.J. and Hoenig, A. (1978) “Mechanical Responses of the Lungs at High Frequencies”, J.Biomechanical Eng (100), 57 (May 1978)

Fung, Y.C. (1993) "Biomechanics - Mechanical Properties of Living Tissues" , p. 474, 2nd ed. (1993) Springer, NY

Gable, T.J. (2000)"Speaker Verification Using Acoustic and Glottal Electromagnetic Micro-power Sensor (GEMS) Data" Thesis, Dec. 2000. University of California at Davis, available through ProQuest Digital Dissertations, index number 9997362

Gabriel, S., Lau, R.W., and Gabriel, C., (1996)“The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum of tissues” Phys. Mod. Biol. **41**, 2271-2293 (1996) IOP Publishing Ltd., UK

Holzrichter, J.F. (1995) “New Ideas for Speech Recognition and Related Technologies”, Lawrence Livermore National Laboratory Report, UCRL-UR-120310 , 1995 . Available from Lawrence Livermore National Laboratory library, or from NTIS in Springfield, VA at http://www.ntis.gov/ordering.htm

Holzrichter, J.F., Burnett, G.C., Ng, L.C., and Lea,W.A. (1998) "Speech Articulator Measurements Using Low Power EM Wave Sensor" , J. Acoust. Soc. Am. **103** (1) 622, 1998. Also see the Website http://speech.llnl.gov/

Holzrichter, J.F and Burnett, G.C. (1999) "Human Speech Articulator measurements using low power, 2 GHz Homodyne Sensors" 24th international conference on infrared and millimeter waves", Lombardo. L. A. ed, Monterey, CA. Sept 5 (1999), Kluwer Publishing; see also Lawrence Livermore Laboratory report UCRL-JC-134775

Holzrichter, J.F., Kobler, J. B., Rosowski, J.J., Burke, G.J., "EM wave simulations and measurements of glottal structure dynamics – a Technical Report" UC Report UCRL-JC-147775 , 2003. Available on the Website http://speech.llnl.gov/ or from Lawrence Livermore National Laboratory library, or from NTIS in Springfield, VA at http://www.ntis.gov/ordering.htm

Ishizaka, K., Matsudaira, M., Kaneko, T., (1976) "Input acoustic-impedance measurement of the subglottal system" J. Acoust. Soc. Am., **60** (1), 190 (1976)

Ishizaka, K., French, J.C., and Flanagan, J.L., (1975) "Direct Determination of Vocal Tract Wall Impedance," IEEE Trans. Acoustics, Speech, and Signal Processing **ASSP-11** (4), 370 (1975)

McEwan, T.E., (1994) U.S. Patent Nos. 5,345,471, 5,361,070, and 5,573,012 .

Miller, E.K., Medgyesi-Mitschang, L, and Newman, E.H. , eds (1992) "Computational Electromagnetics-Frequency-Domain Method of Moments" IEEE Press, New York, 1992 .

Ng, L. C.; Burnett, G. C.; Holzrichter, J. F.; and Gable, T. J. (2000) "De-noising of Human Speech Using Combined Acoustic and EM Sensor Signal Processing", Icassp-2000, Istanbul, Turkey, June 6, 2000

Roark, R.J. and Young, W.C., (1975) "Formulas for Stress and Strain", 5th ed., 1975, McGraw-Hill Book Co., NY

Rosowski, J.J.; Ravicz, M.E.; Teoh, S.W.; and Flandermeyer, D., (1999) "Measurements of Middle-Ear Function in the Mongolian Gerbil, a Specialized Mammalian Ear", Audiol Neurootol 1999;**4**:129-136 (1999)

Rothenberg, M. (1981) "Some relations between glottal flow and vocal fold contact area," ASHA Rep. **11**, 88-96 .

Sharpe, R.M., J.B. Grant, N.J. Champagne, W.A. Johnson, R.E. Jorgenson, D.R. Wilton, W.J. Brown, and J.W. Rockway, (1997) "EIGER: Electromagnetic Interactions GeneRalized" IEEE AP-S International Symposium and North American URSI Radio Science Meeting, Montreal, Canada, July 1997, pp. 2366--2369.

Skolnik, M. (1990). "*Radar Handbook*," 2nd edition., McGraw-Hill, New York

Stevens, K.N. (2000) "Acoustic Phonetics" MIT Press, Cambridge, MA

Svirsky, M.A., Stevens, K.N., Matthies, M.L., Manzella, J. Perkell, J.S., and Wilhelms-Tricarico, R. (1997) "Tongue surface displacement during bilabial stops", J.Acoust. Soc. Am. **102** (1), 562 (1997)

Titze, I. R.(1984). "Parameterization of the glottal area, glottal flow, and vocal fold area" J. Acoust. Soc. Am. **74**(2), 570-580

Titze, I.R., (1990) "Interpretation of the Electroglottographic Signal" , J. of Voice 4 (1), 1-9 (1990) Raven Press, Ltd. NY

Titze, I.R., (1994)"Principles of Voice Production" Prentice Hall, NJ, 1994

Titze, I.R., Story, B.H., Burnett, G.C., Holzrichter, J.F., Ng, L.C., Lea, W.A., (2000) "Comparison between electroglottography and electromagnetic glottography" **107** (1), 581 (2000)